Supporting Information

Lee and Noppeney 10.1073/pnas.1115267108

SI Experimental Procedures

Subjects. In the musicians group (M+ group; mean ± SEM age = 24.3 ± 5.3 y), 18 subjects were amateur pianists who started to learn to play the piano at an average of 7.9 y (SEM = 1.9 y) and have been playing the piano for an average of 16.4 y (SEM = 5.5 y). Furthermore, during the last 3 y, they have been practicing on the piano for an average of 3.33 h/wk (SEM = 1.64 h/wk). In the nonmusicians group (M− group; mean ± SEM age = 27.1 ± 3.4 y), 3 subjects had basic music training (<3 mo; one learned flute, one learned drums, and one learned drums and bass guitar), and 16 subjects had no music training at all. All were German native speakers who participated in both psychophysical and functional MRI (fMRI) study after giving informed consent.

Five additional subjects participated only in the psychophysical study and therefore, were not included in this report. The study was approved by the joint human research review committee of the Max Planck Society and the University of Tübingen.

Stimuli. Stimulus material was taken from close-up audiovisual recordings of (*i*) a female actress' face looking straight into the camera, uttering short sentences, or (*ii*) one male right hand playing short piano melodies on the keyboard (one octave of piano keys was included). Audio and video was recorded with a camcorder (HVX 200 P; Panasonic). The video was acquired at 25 frames/s phase alternation line (PAL = 768 × 567 pixels); audio was acquired at 48 kHz (two channels). Sentences were short, neutral statements in German (four to five words and seven to eight syllables) (*SI Results*, *Appendix A*). The piano melodies were generated to match the rhythm and number of syllables of the speech sentences. The visual and audio recordings were then digitized into MPEG-4 (H.264) format files. Using Adobe Premier Pro (Adobe Systems), the visual file was cropped into one single complete visual stimulus (speech or music) that was preceded and followed by 15 frames of (*i*) neutral facial expression (speech) or (*ii*) a still hand image (music).

Experimental Setup and Stimulus Presentation. The audiovisual stimulus onset asynchrony (AV-SOA) of the separate audio and video files was manipulated with Psychophysics Toolbox version 3, which runs on Matlab 2008 (MathWorks Inc) and a Macintosh laptop OS-X 10.5.6 (Apple).

Psychophysical study. Visual stimuli (size = 8.89° × 7° visual angle) were projected using a CRT monitor (Sony), and the subjects' heads were stabilized using a chin rest. Auditory stimuli were presented at ∼75 dB sound pressure level (SPL) through headphones.

fMRI study. Visual stimuli (size = 13.5° × 10° visual angle) were back-projected onto a Plexiglas screen using an LCD projector (JVC Ltd.) visible to the subject through a mirror mounted on the MR head coil. Auditory stimuli were presented at ∼75 dB SPL using MR-compatible headphones (MR Confon).

fMRI Data Acquisition. A Siemens TRIO TIM 3T MRI scanner (Siemens) was used to acquire both T1 structural volume images [repetition time (TR)/echo time (TE)/inversion time (TI) = 2,300/9.38/1,100 ms, 176 slices, matrix = 256 × 240, spatial resolution = 1 × 1 × 1 mm^3 voxels] and T2*-weighted axial echoplanar images with blood oxygenation level-dependent contrast (gradient echo, TR/TE = 3,000/40 ms, 42 axial slices acquired in ascending direction, matrix = 64 × 64, slice thickness = 3 mm, interslice gap = 0.3 mm, spatial resolution = 3 × 3 × 3.3 mm^3 voxels). There were six sessions with a total of 224 volume images per session.

Behavioral Analysis: Psychophysics Study (Before fMRI Study). For each subject, the proportion of synchronous responses was computed for each of the 13 AV-SOA levels separately for the music and speech. To refrain from making any distributional assumptions, the psychometric function was estimated for the proportion synchronous responses using local quadratic fitting as a nonparametric approach (1). The bandwidth for the local quadratic fitting was optimized individually for each subject in a cross-validation procedure. The audiovisual temporal integration window was estimated as the integral of the fitted psychometric function bounded by ±360 ms and entered into a mixed design ANOVA at the random effects level. The mixed design ANOVA for the temporal integration window with stimulus class (speech and music) as within-subject factor and group (nonmusicians and musicians) as between-subject factor was performed using the SPSS program (version 15.0), and results were reported after Greenhouse Geisser correction. Furthermore, results of the posthoc *t* tests were corrected for multiple comparisons using the Holm–Bonferroni method (2).

fMRI—Dynamic Causal Modelling. Dynamic causal modelling (DCM) treats the brain as a dynamic input state–output system (3). The inputs correspond to conventional stimulus functions encoding experimental manipulations. The state variables are neuronal activities, and the outputs are the regional hemodynamic responses measured with fMRI. The idea is to model changes in the states, which cannot be observed directly, using the known inputs and outputs. Critically, changes in the states of one region depend on the states (i.e., activity) of others. This dependency is parameterized by effective connectivity. There are three types of parameters in a DCM: (*i*) input parameters that describe how much brain regions respond to experimental stimuli, (*ii*) intrinsic parameters that characterize effective connectivity among regions, and (*iii*) modulatory parameters that characterize changes in effective connectivity caused by experimental manipulation. This third set of parameters, the modulatory effects, allows us to explain fMRI asynchrony effects by changes in coupling among brain areas. Importantly, this coupling (effective connectivity) is expressed at the level of neuronal states. DCM employs a forward model, relating neuronal activity to fMRI data that can be inverted during the model fitting process. Put simply, the forward model is used to predict outputs using the inputs. The parameters are adjusted (using gradient descent) so that the predicted and observed outputs match under complexity constraints. This adjustment corresponds to the model fitting.

SI Results

Eye Movement Monitoring. Horizontal and vertical eye movements were recorded in 19 subjects (10 nonmusicians and 9 musicians) inside the scanner using Eye-Trac series long-range optics for fMRI (Applied Science Laboratories) to obtain and compare subjects' points of fixation on the computer screen across conditions (60-Hz sampling rate; tracking resolution = 0.1°).

Eye recordings were calibrated with standard eccentricities between 4° and 15° to determine the deviation from the fixation cross (fixation position was posthoc offset corrected). Eye position data were automatically corrected for blinks. For each trial, the mean distance (degree) from the fixation cross and the number of saccades (defined by eye velocity threshold >2°/s) were quantified. Subjects consistently fixated (mean distance from fixation cross ∼1.1°). Both of these indices (i.e., mean distance from fixation point and number of saccades) were entered into a repeated measures ANOVA with stimulus class (speech and music) and audiovisual (a)synchrony (synchronous and asynchronous) as within-subjects variables and group (nonmusicians and musicians) as the between-subject variable. Across subjects, these analyses did not show any significant effects of conditions. Specifically, the ANOVA for mean distance from fixation cross did not reveal a significant main effect of group [$F_{(1,17)} < 1$, non significant (n.s.)], stimulus class [$F_{(1.0,17.0)} < 1$, n.s.], or audiovisual (a)synchrony [$F_{(1.0,17.0)} < 1$, n.s.]. Similarly, the ANOVA for mean number of saccades did not reveal a significant main effect of group [$F_{(1,17)} < 1$, n.s.], stimulus class [$F_{(1.0,17.0)} < 1$, n.s.], or audiovisual (a)synchrony [$F_{(1.0,17.0)} < 1$, n.s.]. These findings show that the subjects' eye movements did not differ across conditions and groups. Therefore, it is unlikely that eye movements can account for the activation differences observed in fMRI data.

Appendix A.

i) Piraten rauben Schiffe aus.
ii) Pollen fliegen durch die Luft.
iii) Täter meiden ihre Opfer.
iv) Tische haben vier Beine.
v) Der Postbote bringt die Post.
vi) Die Ampel regelt den Verkehr.
vii) Katzen haben neun Leben.
viii) Kugeln haben eine runde Form.
ix) Gänse haben weisse Federn.
x) Geister machen vielen Menschen Angst.
xi) Inder haben dunkle Haut.
xii) Otter haben weiches Fell.
xiii) Bären haben grosse Tatzen.
xiv) Blumen haben bunte Blüten.
xv) Nachspeisen enthalten oft viel Fett.
xvi) Nonnen tragen schwarze Kutten.

1. Zychaluk K, Foster DH (2009) Model-free estimation of the psychometric function. *Atten Percept Psychophys* 71:1414–1425.
2. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6: 65–70.
3. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19: 1273–1302.

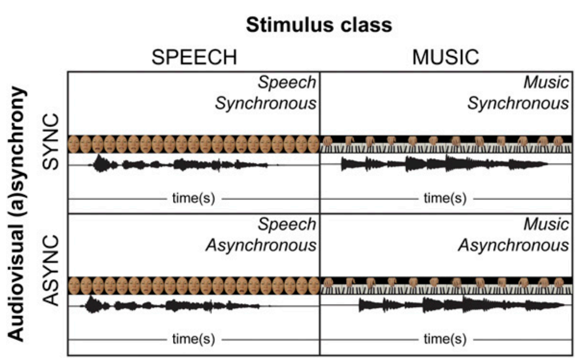

Fig. S1. The experimental paradigm. The 2 × 2 factorial design manipulated stimulus class (speech and music) and audiovisual (a)synchrony (synchronous and asynchronous).

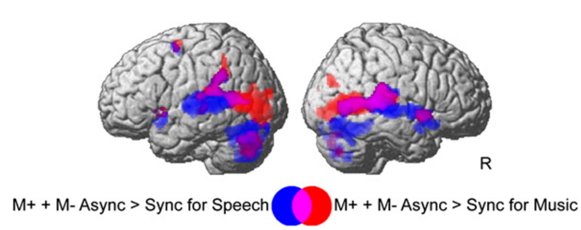

Fig. S2. Asynchrony effects averaged across nonmusicians (M−) and musicians (M+) for speech (blue), music (red), and common for speech and music (magenta) are rendered on a template brain. Height threshold: $P < 0.001$, uncorrected and extent threshold > 20 voxels; inclusively masked with the search mask for illustration purposes.

Table S1. Asynchrony effects for speech averaged across groups

Brain region	Cluster size	MNI coordinates x	y	z	z Score (peak)	P value	z Score (peak) Speech M+	Speech M−	Music M+	Music M−
M+ + M− async > sync for speech										
R. posterior superior temporal sulcus/middle temporal gyrus	3,436	54	−40	6	5.2	0.001	3.9	3.7	4.4	1.9
R. posterior superior temporal sulcus/middle temporal gyrus		58	−26	2	5.2	0.001	4.2	3.3	2.5	1.5
R. posterior superior temporal sulcus/middle temporal gyrus		54	−24	−2	5.1	0.001	4.1	3.3	2.5	1.3
R. extrastriate cortex		46	−58	4	4.4	0.04	3.0	3.2	4.6	2.7
R. superior temporal pole		52	16	−6	4.4	0.04	2.3	3.9	3.8	1.0
R. cerebellum (VI)		40	−58	−22	4.3	0.05	3.7	2.4	2.6	0.6
L. posterior superior temporal sulcus/middle temporal gyrus	2,157	−52	−44	8	4.7	0.01	3.4	3.4	1.3	1.4
L. posterior superior temporal sulcus/middle temporal gyrus		−54	−42	4	4.7	0.01	3.1	3.6	1.1	0.7
L. posterior superior temporal sulcus/middle temporal gyrus		−48	−42	10	4.5	0.02	3.2	3.3	1.2	1.4
L. posterior superior temporal sulcus/middle temporal gyrus		−48	−32	6	4.4	0.04	3.1	3.2	1.4	1.7
L. supramarginal gyrus		−58	−44	30	4.6	0.01	2.9	3.7	3.3	2.5
L. extrastriate cortex		−50	−54	6	4.3	0.05	3.5	2.7	3.3	1.7
L. anterior middle temporal gyrus		−62	−26	0	4.5	0.02	3.1	3.4	1.2	0.4
L. superior temporal pole	71	−52	12	−8	4.4	0.04	2.5	3.7	2.9	1.5
L. superior temporal pole		−50	16	−10	4.3	0.04	2.5	3.7	2.7	0.7
L. cerebellum (Crus II/VIIb)	2,001	−14	−76	−38	5.1	0.002	3.3	4.0	3.7	1.4
L. cerebellum (Crus II/VIIb)		−16	−72	−44	5.0	0.002	3.7	3.6	3.6	1.5
L. cerebellum (Crus II/VIIb)		−32	−70	−28	4.7	0.007	2.8	4.0	2.9	0.5
L. cerebellum (Crus II/VIIb)		−20	−68	−48	4.7	0.01	3.4	3.3	1.8	1.3
L. cerebellum (Crus II/VIIb)		−36	−72	−26	4.5	0.02	2.7	3.8	1.7	0.5

P value, Corrected at peak level for multiple comparisons within the search volume of interest (see *Experimental Procedures, Search Volume Constraints*).

Table S2. Asynchrony effects for music averaged across groups

Brain region	Cluster size	MNI coordinates x	y	z	z Score (peak)	P value	z Score (peak) Speech M+	Speech M−	Music M+	Music M−
M+ + M− async > sync for music										
R. extrastriate cortex	2,073	46	−56	6	5.0	0.002	2.7	3.1	4.7	2.6
R. extrastriate cortex		44	−70	2	4.6	0.02	3.0	2.7	3.3	3.3
R. extrastriate cortex		56	−64	8	4.3	0.04	1.7	3.1	3.7	2.5
R. posterior middle temporal gyrus		54	−44	8	4.7	0.008	3.1	3.7	4.4	2.4
R. posterior superior temporal sulcus/gyrus		56	−40	8	4.6	0.01	3.8	3.5	4.5	2.1
R. posterior superior temporal sulcus/gyrus		56	−48	8	4.6	0.01	2.4	3.4	4.5	2.2
L. extrastriate cortex	1,765	−50	−68	8	5.1	0.001	2.0	2.1	3.8	3.6
L. posterior superior temporal gyrus		−64	−40	18	4.6	0.02	2.1	2.5	3.7	2.9
L. supramarginal gyrus		−58	−46	36	4.3	0.04	1.6	3.9	3.9	2.3
L. inferior parietal lobule		−56	−46	40	4.6	0.01	0.9	3.0	4.4	2.2
L. cerebellum (Crus II/VIIb)	542	−12	−80	−32	4.5	0.02	2.0	3.5	4.5	2.0
L. cerebellum (Crus II/VIIb)		−16	−72	−40	4.4	0.03	3.4	3.1	4.5	1.9
L. cerebellum (Crus II/VIIb)		−14	−74	−26	4.3	0.05	1.8	3.0	4.1	2.0

P value, Corrected at peak level for multiple comparisons within the search volume of interest (see *Experimental Procedures, Search Volume Constraints*).

Table S3. Asynchrony effects for speech and music that are common for nonmusicians and musicians

		Nonmusicians (M−)					Musicians (M+)				
		MNI coordinates					MNI coordinates				
Brain regions	Cluster size	*x*	*y*	*z*	*z* Score (peak)	*P* value	*x*	*y*	*z*	*z* Score (peak)	*P* value
Asynchrony effects for speech that are common for nonmusicians and musicians		M− async > sync for speech					M+ async > sync for speech				
R. mid superior temporal sulcus/gyrus	180	54	−42	6	3.8	0.02	58	−26	4	4.2	0.003
R. mid superior temporal sulcus/gyrus		58	−28	2	3.5	0.04	56	−24	−2	4.2	0.004
							54	−38	8	3.9	0.01
L. posterior superior temporal sulcus/gyrus	36	−54	−42	4	3.6	0.03	−52	−44	8	3.4	0.078
L. cerebellum (Crus II/VIIb)	61	−14	−76	−38	4.0	0.007	−16	−72	−44	3.7	0.03
L. cerebellum (Crus II/VIIb)		−16	−74	−44	4.0	0.007	−14	−72	−38	3.6	0.04
L. cerebellum (Crus II/VIIb)		−12	−76	−42	3.7	0.02					
Asynchrony effects for music that are common for nonmusicians and musicians		M− async > sync for music					M+ async > sync for music				
L. extrastriate cortex	84	−50	−64	6	3.7	0.05	−52	−70	6	4.0	0.001

P value, Corrected at peak level for multiple comparisons within the search volume of interest (see *Experimental Procedures, Search Volume Constraints*).